INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Regularization and Scale Space

Mads Nielsen, Luc Florack, and Rachid Deriche

N° 2352

Septembre 1994

PROGRAMME 4

Robotique,

image

et vision

Rapport de recherche

1994

ISSN 0249-6399

Regularization and Scale Space

Mads Nielsen, Luc Florack, and Rachid Deriche

Programme 4 — Robotique, image et vision
Projet ROBOTVIS

Rapport de recherche n°2352 — Septembre 1994 — 40 pages

Abstract: Computational vision often needs to deal with derivatives of digital images. Derivatives are not intrinsic properties of a digital image; a paradigm is required to make them well-defined. Normally, a linear filtering is applied. This can be formulated in terms of scale space, functional minimization or edge detection filters. In this paper, we take regularization (or functional minimization) as a starting point, and show that it boils down to a ordered set of linear filters of which the Gaussian is the first if we require the semi group constraint to be fulfilled. This regularization implies the minimization of a functional which contains terms up to infinite order of differentiation. If the functional is truncated at second order, the Canny-Deriche filter arises. Furthermore, we show that the nth order Canny-optimal edge detection filter implements nth order regularization. We also show, that higher dimensional regularization in its most general form boils down to a rotation of the one dimensional case, when Cartesian invariance is imposed. This means that results from 1D regularization are easily generalized to higher dimensions. Finally, we show that regularization in its most general form can be implemented as recursive filtering without any approximation.

Key-words: Regularization, scale space, semi-group constraint, Cartesian invariance, well-posed differentiation, edge-detection filters.

(Résumé : tsvp)

This work was supported by the Danish Council of Technical Sciences, the Danish Research Academy and ERCIM

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Régularisation et Espace Echelle

Résumé : L'estimation de grandeurs différentielles à partir d'images est un problème important en Vision par Ordinateur. Ceci nécessite en particulier d'estimer correctement les dérivées dans les images et il est bien connu que l'estimation de ces grandeurs, qui ne sont pas des grandeurs intrinsèques aux images, est un problème mal posé qui nécessite d'être régularisé et bien défini. A cette fin, un filtrage linéaire est généralement appliqué. Ceci est formulé en terme d'espace échelle, de minimisation de fonctionnelles ou de filtre de détection de contours. Dans cet article, nous montrons que l'approche régularisation ou minimisation de fonctionnelle permet de dériver un ensemble ordonné de filtres linéaires parmi lesquels seul le filtre Gaussien respecte la contrainte de semi-groupe. Ceci implique la minimisation d'une fonctionnelle qui contient des termes de différentiation d'ordre infinis. Nous montrons alors que le filtre de Canny-Deriche correspond à la solution particulière où la fonctionnelle à minimiser est du second ordre. Nous montrons en plus que l'opérateur optimal de Canny du n-ème ordre correspond à la minimisation d'une fonctionnelle qui contient des termes d'ordre n. La généralisation au cas multi-dimensionnel est aussi considérée. Finalement nous montrons qu'une opération de régularisation dans sa forme la plus générale peut être mise en oeuvre au travers de filtres récursifs sans aucune approximation.

Mots-clé : Régularisation, Espace échelle, Contrainte de semi-groupe, Invariance Cartésienne, Différentiation, Opérateurs de détection de contours.

1 Introduction

Given a digital signal in one or more dimensions, we want to define its derivatives in a well-posed way. This can be done in a distributional sense [1], by convolving the signal with a smooth test function. Instead of taking the derivatives of the signal, we take the derivatives of the convolved signal by deriving the smooth filter prior to convolution. In this way derivatives of any integrable signal are operationally defined and well-posed.

Distributional differentiation has been implemented in various conceptually different ways. In computational vision Gaussian scale space, regularization or edge detection filters are typically applied. In the following we briefly describe these three methods and their motivation and discuss their relations.

Gaussian scale space [2] can be motivated from different points of views. Koenderink [3] has introduced it as a one parameter family of blurred images which fulfill the causality criterion: every isophote in scale space should be upwards convex. This requirement expresses, in a precise sense, that coarse scale details should have a cause at finer scales. It essentially singles out the normalized Gaussian filter. Florack [4] defines a visual front-end to be linear, spatially isotropic, spatially homogeneous and scale invariant, leading to the Gaussian as well. Because the Gaussian is smooth, all derivatives of the images are well-defined in Gaussian scale space.

In Tikhonov regularization theory [5, 6], we look for a differentiable function, which in some sense is closest to our signal. The measure of difference between the solution and the signal is a functional of the solution. In this way the problem of well-posedness of differentiation is turned into a functional minimization problem. Using regularization we can specify up to which order we want the solution to be differentiable.

In edge detection many methods have been used. One method is to define a sample edge as well as criteria for an optimal edge detection. Canny [7] used a step edge and wanted to find the optimal linear detection filter according to criteria of signal-to-noise-ratio, localization and uniqueness of detection. Deriche [8] used these criteria on an infinite domain to find an optimal filter.

In the following section we review the regularization and show how it relates to Gaussian scale space. Furthermore, we show that the Canny-optimal Deriche-filter corresponds to a regularization, and we generalize the Canny optimality to any order and show that it corresponds to a regularization as well. Finally, we show that regularization can be implemented in an efficient way using recursive filtering.

2 Linear Regularization

Regularization of a signal $g \in \mathcal{L}^2(\mathbb{R})$ can be formulated [5] as the minimization with respect to the regularized solution f of an energy functional.

Definition 1 *The Tikhonov regularized solution f of the signal $g \in \mathcal{L}^2(\mathbb{R})$ is the one that minimizes the energy functional*

$$E[f] \equiv \frac{1}{2}\int dx \left((f-g)^2 + \sum_{i=1}^{\infty} \lambda_i (\frac{\partial^i}{\partial x^i} f)^2\right) \quad , \tag{1}$$

with nonnegative λ_i.

When $\lambda_n \neq 0$ and $\lambda_i = 0$ for all $i > n$ we talk about nth order regularization. The solution can be found by linear filtering.

Proposition 1 *Linear convolution of the signal $g \in \mathcal{L}^2(\mathbb{R})$ by the filter h, having the Fourier transform*

$$\hat{h} = \frac{1}{\sum_{i=0}^{\infty} \lambda_i \omega^{2i}} \quad , \tag{2}$$

yields the solution of the regularization of definition 1. By definition $\lambda_0 \equiv 1$.

Proof In Fourier domain the energy functional (1) yields according to Parsevals Theorem

$$E[\hat{f}] = \frac{1}{2}\int d\omega \left((\hat{f}-\hat{g})^2 + \sum_{i=1}^{\infty} \lambda_i \omega^{2i} \hat{f}^2\right) \quad .$$

A necessary condition for E to be minimized with respect to $\hat{f}$ is that the variation of E is zero:

$$\begin{aligned} 0 &= \frac{\delta E}{\delta \hat{f}} = (\hat{f}-\hat{g}) + \sum_{i=1}^{\infty} \lambda_i \omega^{2i} \hat{f} \\ &\Leftrightarrow \\ \hat{f} &= \frac{1}{1+\sum_{i=1}^{\infty} \lambda_i \omega^{2i}}\, \hat{g} \quad . \end{aligned}$$

The optimal $\hat{f}$ can thus be found as a linear filtering of the initial signal g with a filter defined by the constants λ_i. Defining $\lambda_0 \equiv 1$ we obtain the linear filter h given by (2).

□

This means that any regularization using only sums of quadratic terms of the derivatives of the solution can be reformulated as a linear convolution with the Fourier inverse (provided it exists) of $\hat{h}$ as given by (2). In practice, the derivation of a regularized signal, can be performed during the convolution, by convolving with the derivative of the filter.

2.1 Projection space

Regularization can be formulated as a "projection" of the initial function into a function space, in which all functions have the desired regularity properties. In Tikhonov regularization a square integrable function is mapped to a function within a Sobolev space of specified order. A Sobolev space of order N is the space of all functions which are square integrable, and have derivatives up to order N which are all well-defined and square integrable[1].

Sobolev spaces have the property of containing all other Sobolev spaces of higher order:

$$S^0 \supset S^1 \supset S^2 \supset \ldots \supset S^n \quad .$$

The mapping performed by Tikhonov regularization is normally a mapping from a zero order Sobolev space S^0 into a Sobolev space of higher order. In general one may have the regularization $T : S^a \mapsto S^b$, where $a \leq b$. In the case where the initial element is in S^b, the regularization is not the identity operation, because it will in this case map to a Sobolev space of even higher order. Because a projection per definition is idempotent (i.e. $T^2 = T$), we cannot talk about regularization as a projection. We make the following proposition:

Proposition 2 *The nth order regularization maps a function $g \in S^a$, $g \notin S^{a+1}$ into $f \in S^{a+n}$, $f \notin S^{a+n+1}$.*

[1]Definitions of Sobolev spaces using non-integer order differentiation exists. Furthermore, Sobolev spaces are often characterized by one more index, which refers the norm being used (i.e. 2 in our case of square integrability).

Proof We argue in Fourier space. Because $g \in S^a$, its Fourier transform and its derivatives up to ath order are well-defined and square integrable. Furthermore, the regularization filter h_n, which performs nth order regularization, has the same properties up to order n. We see, that the regularized signal f in Fourier space can be written as

$$\hat{f} = \hat{h}_n \cdot \hat{g} \quad .$$

All derivatives up to order $(a+n)$ of this are well-defined and square integrable, because the product of two square integrable functions is square integrable:

$$\hat{f}(i\omega)^j = \hat{h}_n(i\omega)^k \cdot \hat{g}(i\omega)^l \quad ,$$

where j=k+l. For $j \leq a+n$, we can find (k,l) so that $k \leq n$ and $l \leq a$, but not for $j > a+n$.

□

2.2 Semi-Group Constraint

We now add constraints on the filter h stating that g should not play a privileged role, but is to be treated on equal footing with any reconstructed signal f. This constraint is most intuitive when the signal g is given as a digitized signal. In this case the digitizer has already performed a filtering of g. If g then plays a privileged role, it is determined by the digitizer. For further argumentation see Florack [4] or Lindeberg [9]. If we embed h into a 1-parameter family $h(s)$, we may formulate this criterion as a *semi-group* property of $h(s)$.

Definition 2 *The semi-group property of the 1-parameter family of filters $h(s)$ is*

$$\forall s, t \in I\!R_+ : \; h(s \oplus_p t) = h(s) * h(t) \quad ,$$

where the parameter-concatenation is the p-norm addition

$$s \oplus_p t = (s^p + t^p)^{1/p} \quad .$$

We call the filter-parameter the *scale parameter*, because it resembles the scale parameter in common scale space theory.

Proposition 3 *If we associate a dimension of length to the parameter $s \in I\!R^+$ of h, then the regularization of a signal g fulfilling the semi-group constraint using the p-norm addition implies filtering with a filter from the 1-parameter class of filters*

$$\hat{h}_p(\omega, t) = \frac{1}{\sum_{i=0}^{\infty} \frac{t^{ip}}{ip!} \omega^{2ip}} = e^{-(\omega^2 t)^p} \quad ,$$

here given in the Fourier domain.

Proof By dimensional analysis of (2) we find that $[\lambda_k] = [\lambda_1]^k \equiv [s]^k$. The power of s is chosen to resemble scale space theory but any other remapping of $s \mapsto s^m$ could be used as well, and would yield the same filter, but with a redefined scale parameter. We conclude that we can write

$$\lambda_k(s) = \frac{1}{k!}(a_k s)^k$$

where a_k is a set of dimensionless parameters. This is the most general form of $\lambda_k(s)$, when we assume, that there exist no other dimensionfull parameter than s. This dimensional argument removed many possible filters in the class of filters fulfilling the semi-group constraint [11]. Applying the semi-group property to the actual form of h given by (2) we find this to be identical to

$$\forall s, t \in \mathrm{I\!R}_+ : \ \sum_{i=0}^{\infty} \lambda_i(s \oplus_p t)\omega^{2i} = \left(\sum_{i=0}^{\infty} \lambda_i(s)\omega^{2i}\right)\left(\sum_{i=0}^{\infty} \lambda_i(t)\omega^{2i}\right) \quad .$$

By separating into terms of equal power of ω, we find

$$\forall s, t \in \mathrm{I\!R}_+^2 : \ \lambda_k(s \oplus_p t) = \sum_{l=0}^{k} \lambda_l(s)\lambda_{k-l}(t) \quad .$$

For clarity we now analyse the situation of $p = 1$, afterwards we deal with the general case. The lowest order terms of this, using the 1-norm scale-concatenation, read as follows (recall $\lambda_0(s) \equiv 1$)

$$\begin{aligned}
\lambda_0(s+t) &= 1 \\
\lambda_1(s+t) &= \lambda_1(s) + \lambda_1(t) \\
\lambda_2(s+t) &= \lambda_2(s) + \lambda_2(t) + \lambda_1(s)\lambda_1(t)
\end{aligned}$$

Let us choose $a_1 = \tau$. By dimensional analysis and induction we then find $a_k = \tau$ independent of k, resulting in a filter of the form $(t = \tau s)$

$$\hat{h}(\omega, t) = \frac{1}{\sum_{i=0}^{\infty} \frac{t^i}{i!}\omega^{2i}} = e^{-\omega^2 t} \quad ,$$

which is just the well-known Gaussian. We formulate this as the result:

Result 1 *The semi-group property, when applied to Tikhonov regularization, yields Gaussian scale space if the 1-norm addition is used for concatenation of scales, and a dimension*[2] *of length is associated to s.*

Applying the general p-norm scale-concatenation results in a broader class of filters. We find

$$a_1(s^p + t^p)^{1/p} = a_1 s + a_1 t \quad .$$

This is only true either if $p = 1$ as in the above calculations, or if a_1 is zero. If all lower order a_i are zero, we find

$$a_m(s^p + t^p)^{m/p} = a_m s^m + a_m t^m \quad .$$

This yields $a_m = 0$ unless $m = p$. This shows that only for $p \in \mathbb{N}$, we find non-trivial solutions. For $p \in \mathbb{N}$, we find the same coefficients as above, when mapping $i \mapsto ip$. We find the filters

$$\hat{h}_p(\omega, t) = \frac{1}{\sum_{i=0}^{\infty} \frac{t^{ip}}{ip!}\omega^{2ip}} = e^{-(\omega^2 t)^p} \quad .$$

This is a class of filters all having the semi-group property for different scale-concatenation norms p.

□

All $\hat{h}_p$ are Fourier-invertible. For $p = 1$ it is the Gaussian, for p being increased towards infinity it converges towards the ideal low-pass filter with cut-off frequency $\omega_c = 1/\sqrt{t}$.

[2]Without the dimensional argument, a free parameter α_k exists for each order of regularization. This implies a much broader, but physically meaningless, class of filters.

In Florack [4] the criteria of isotropy and separability are used in addition to the semi-group property to single out the Gaussian as the unique scale space filter. In terms of Tikhonov regularization, we can use another argumentation to single out the Gaussian: We require the derivatives of the solutions to be equally weighted in a Taylor series sense, and get the following result:

Result 2 *The only linear estimation scheme, which has an equal weighting of the different derivatives of the solution, in a Taylor serries sense, is the Gaussian scale space extension.*

We can also reverse the argumentation: Gaussian scale space can be implemented as the minimization of an energy functional (using a smoothness term of infinite order). The corresponding energy function is

$$E(f) = \frac{1}{2}\int dx (f-g)^2 + \sum_{i=1}^{\infty} \frac{t^i}{i!}[\frac{\partial^i}{\partial x^i} f]^2 \quad .$$

The Euler-Lagrange equation of this is

$$0 = (f-g) + \sum_{i=1}^{\infty} \frac{(-t)^i}{i!} \frac{\partial^{2i}}{\partial x^{2i}} f = e^{-t\triangle} f - g \quad ,$$

where the exponential function is defined by its Taylor series expansion. This differential equation is an equation using only spatial derivatives, but having the same solutions as the Heat Equation. The solution can formally be written as

$$f = e^{t\triangle} g \quad .$$

By differentiation with respect to t, we obtain the Heat Equation:

$$\begin{cases} \frac{\partial}{\partial t} f &= \triangle f \\ f|_{t=0} &= g \end{cases} \quad .$$

The same derivation can be made for the other semi-group filters under the mapping $\triangle \mapsto \triangle^p$.

2.3 Boundary conditions

The regularization results in an Euler-Lagrange equation, which is a differential equation of order $2N$, where N is the order of regularization. The Euler-Lagrange equation can be solved, but the solution is only determined up to a set of boundary conditions. These boundary conditions can be determined as the boundary conditions, which minimize the energy functional. When we minimize the energy functional not just according to the differential structure of the solution, but also according to the solution on the boundary $\mathcal{B}$, we find the so-called *natural boundary conditions* or *free boundary conditions*:

$$\forall j \in \mathbb{N}_0,\ \forall x \in \mathcal{B}:\ \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial x}\right)^{n-1} E_{f^{(j+n)}} = 0 \quad ,$$

where $E_{f^{(n)}}$ denotes the integrand of the energy functional (symbolically) differentiated with respect to the nth derivative of f. These boundary conditions does not in general imply that all derivatives up to some order have to be zero (as often stated in the literature).

In the case of Tikhonov regularization (1), we find the boundary conditions:

$$\forall j \in \mathbb{N}_0,\ \forall x \in \mathcal{B}:\ \sum_{n=1}^{\infty} (-)^{n+j-1} \lambda_{n+j} \left(\frac{\partial}{\partial x}\right)^{2n+j-1} f = 0 \quad .$$

In the case of regularization implementing the Heat Equation, this system of equations is invertible, and we find

$$\forall n \in \mathbb{N},\ \forall x \in \mathcal{B}:\ \left(\frac{\partial}{\partial x}\right)^{n} f = 0 \quad .$$

All derivatives of the solution (but not the solution itself) thus have to vanish on the boundary of the domain of the function.

When the Heat Equation is defined on a finite domain, the Gaussian is no longer the corresponding Green's function, and the solution cannot be found by linear filtering with a Gaussian. If Gaussians are used anyhow, we need a transformation into Fourier space (and the assumption of periodicity), an ad hoc cutting off of the linear filters, an extension of the image, or some other method of well-defining the solution. All these methods might violate the principle of causality. The solution can though be found by simulating the Heat Equation, or by minimizing the corresponding functional.

2.4 Scale space interpretation of regularization

We have seen that infinite order regularization with appropriate weights of the different orders boils down to linear scale space. We could reformulate this result as follows: as time evolves, the signal governed by the Heat Equation travels through the minima of a one parameter functional. In general, we would like to be able to construct a partial differential equation which travels through the minima of a given one-parameter functional. We look into the case of Tikhonov regularization. We want to find the function f which minimizes the functional

$$E[f] = \int dx \left((f-g)^2 + \lambda \sum_{i=1}^{n} \lambda_i (\frac{\partial^i}{\partial x^i} f)^2 \right) \quad ,$$

where λ_i are arbitrary. We find the Euler-Lagrange equation by setting the first variation of this to zero

$$(f-g) + \lambda \sum_{i=1}^{n} \lambda_i (-\Delta)^i f = 0 \quad .$$

We perceive f as being a function of space x and scale λ and find by differentiation with respect to λ the following differential equation:

$$(1 + \lambda \sum_{i=1}^{n} \lambda_i (-\nabla)^i) \frac{\partial}{\partial \lambda} f = - \sum_{i=1}^{n} \lambda_i (-\nabla)^i f \quad .$$

This is a differential equation governing f, so that it travels through the minima of the above functional. In Fourier space it yields

$$(1 + \lambda \sum_{i=1}^{n} \omega^{2i}) \frac{\partial}{\partial \lambda} \hat{f} = \sum_{i=1}^{n} \omega^{2i} \hat{f} \quad .$$

In case of first order regularization (i.e. $\lambda_1 = 1$ and all other $\lambda_i = 0$) this is the Heat Equation, but with the following remapping of scale

$$d\lambda = (1 + \lambda \omega^2) ds$$

leading to

$$s(\lambda; \omega) = \frac{\log(1 + \lambda \omega^2)}{\omega^2}$$

In this way first order regularization can be interpreted as Gaussian scale space with a frequency dependent remapping of scale. Note that this remapping of scale is not Fourier invertible, and cannot be expressed as a spatially variyng remapping of scale.

3 Higher dimensional regularization

So far, we have discussed 1-dimensional regularization. In this section we discuss regularization in a D-dimensional space. One might expect a lot of mixed derivatives and products of derivatives of different order to show up in the functional in the most general form. In principle each of these terms might require each an independent weight, leading to a combinatorial explosion of the number of independent parameters in the functional. The purpose of this section is to show that the requirement of invariance of the functional to the choice of Cartesian coordinate system leads to only a single independent parameter for each order of differentiation in the functional.

Assumption 1 *The representation of f and g should be identical in any Cartesian representation, because in general we do not prefer any representation for others.*

In other words, we want the functional $E[f]$ to be an absolute *Cartesian invariant* of f [4]. At this point it is necessary to infer some notation from tensor calculus. We only explain the notation, and not the semantics, for which we refer to Florack [4]. We infer the following notation: $\partial_{\mu_1 \mu_2 \ldots \mu_N}$ is the differential operator $\frac{\partial^N}{\partial x_{\mu_1} \ldots \partial x_{\mu_N}}$, where the μ can be the index of any of the Cartesian coordinates $x_1, \ldots x_D$. We use the Einstein summation convention: We sum over all possibilities of values of the indices repreated twice. Examples of notation in 2D:

$$
\begin{aligned}
\partial_{\mu\nu} f \quad &\equiv f_{\mu\nu} \quad \equiv \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} \\
\partial_{\mu\mu} f \quad &\equiv f_{\mu\mu} \quad \equiv f_{xx} + f_{yy}
\end{aligned}
$$

We do not have to distinguish between covariant and contravariant components, because we work in a Cartesian coordinate system. Nevertheless, some arguments turn out to be clearer if we make the distinction: covariant components have lower indices and contravariant components have upper indices. The metric tensor (in 2D) $\eta_{\mu\nu}$ and its inverse $\eta^{\mu\nu}$ can be used to lower or rise indices (transform contravariant tensors to covariant tensors and vice versa): $\eta^{\mu\nu} f_\mu = f^\nu$. Contravariant and covariant components of the metric tensor is related by the property $\eta_{\mu\nu}\eta^{\nu\rho} = \delta^\rho_\mu$. The symbol δ^j_i is the Kronecker tensor ($\delta^j_i = 1$ if $j = i$ and 0 otherwise). We use paranthesis around indices to indicate a symmetrization:

$$X_{(\mu_1 \ldots \mu_N)} \equiv \frac{1}{N!} \sum_\pi X_{\pi(\mu_1)\ldots\pi(\mu_N)}$$

in which the summation extends over all permutations π of indices $(\mu_1 \ldots \mu_N)$. Finally, by the *local jet* of order n of a function f, we mean the equivalent class of all functions which have the same local derivatives $f_{\mu_1 \ldots \mu_i}$ for all orders $i = 0, \ldots, n$.

Theorem 1 *Any polynomial constructed from products of derivatives $f_{\mu_1 \ldots \mu_i}$, $(i = 0, \ldots, n)$ and metric components $\eta_{\mu\nu}$, $\eta^{\rho\sigma}$, in which all indicies has been contracted is a Cartesian polynomial absolute invariant, and any Cartesian polynomial absolute invariant can be stated in this form.*

The proof is given by Florack [4].

Regularization of N-th order can then be stated as the minimization of the functional:

$$E[f] = \int d\mathbf{x}\, (f - g)^2 + \sum_{i=1}^{2N-1} \sum_{j=1}^{i} \Delta^{\mu_1 \ldots \mu_j}_{\nu_1 \ldots \nu_i} f_{\mu_1 \ldots \mu_j} f^{\nu_1 \ldots \nu_i} \tag{3}$$

where $\Delta^{\mu_1 \ldots \mu_j}_{\nu_1 \ldots \nu_i}$ is a tensor containing the weights of the different terms. This form of functional includes every possible smoothness term which is quadratic in f, and is rotationally invariant. What we will show in the following is, that it only contains N independent parameters. The number of terms in (3) is much larger and dependent on the dimension. All of these terms are in principle mutually independent, but because we integrate over the spatial domain, they become linearly dependent. Here follows the proof. In order to handle this functional and analyse its structure in the case where it is a Cartesian invariant we need the following lemmas:

Lemma 1 *If X and Y vanishes sufficiently fast on the boundary then*

$$\int d\mathbf{x}\, \eta_{\alpha\beta}\eta^{\mu\nu} X_{\mu\nu} Y^{\alpha\beta} = \int d\mathbf{x}\, \delta^{(\mu}_{(\alpha}\delta^{\nu)}_{\beta)} X_{\mu\nu} Y^{\alpha\beta}$$

Proof When X and Y vanishes sufficiently fast on the boundary, we can Fourier transform the first term:

$$\int d\omega \eta_{\alpha\beta}\eta^{\mu\nu}\omega_\mu\omega_\nu\omega^\alpha\omega^\beta XY$$

Because this is symmetric in indices, we can replace $\eta_{\alpha\beta}\eta^{\mu\nu}$ by symmetrisized δ's. Alternatively the same proof could be carried out in the spatial domain by using partial integration.

□

Lemma 2 *Any symmetrized constant tensor $\Delta^{(\nu_1\ldots\nu_i)}_{(\mu_1\ldots\mu_i)}$, which can be expressed as a polynomial in the metric tensor, can be written in the form:*

$$\Delta^{(\mu_1\ldots\mu_i)}_{(\nu_1\ldots\nu_i)} = \sum_{j=0}^{[i/2]} a_j \eta^{(\mu_1\mu_2}\ldots\eta^{\mu_{2j-1}\mu_{2j}}\eta_{(\nu_1\nu_2}\ldots\eta_{\nu_{2j-1}\nu_{2j}}\delta^{\mu_{2j+1}}_{\nu_{2j+1}}\ldots\delta^{\mu_i)}_{\nu_i)}$$

where $[x]$ denotes the integer part of $x \in I\!R$.

Proof When constructing a symmetric tensor as a polynomial in the metric tensor, only two different combinations of a covariant and a contravariant metric tensor yielding a symmetric tensor with free indices can be used as building blocks:

$$\eta_{ab}\eta^{bc} = \delta^c_a \quad \text{and} \quad \eta_{ab}\eta^{cd}$$

When constructing $\Delta^{(\nu_1\ldots\nu_i)}_{(\mu_1\ldots\mu_i)}$ an equal number of covariant and contravariant metric tensors is needed. The essential difference among terms which are symmetrized is then the number of blocks of type one and two. The most general symmetrized polynomial in the metric tensor is then given as linear combinations of terms with different numbers of the two different types of building blocks.

□

Theorem 2 *The integral on the form*

$$I = \int d\mathbf{x}\, \Delta^{\mu_1 \ldots \mu_i}_{\nu_1 \ldots \nu_i} f_{\mu_1 \ldots \mu_i} f^{\nu_1 \ldots \nu_i}$$

where $\Delta^{\mu_1 \ldots \mu_i}_{\nu_1 \ldots \nu_i}$ *is a constant tensor constructed as a polynomial in the metric tensor can be written as*

$$I = \lambda \int d\mathbf{x}\, f_{\mu_1 \ldots \mu_i} f^{\mu_1 \ldots \mu_i} \tag{4}$$

when f *is a real function and is vanishing sufficiently fast at the boundary and* λ *is a constant.*

Proof We find I from the Fourier transform of f as

$$\begin{aligned} I &= \int d\mathbf{x}\, \Delta^{\mu_1 \ldots \mu_i}_{\nu_1 \ldots \nu_i} f_{\mu_1 \ldots \mu_i} f^{\nu_1 \ldots \nu_i} \\ &= \Delta^{\mu_1 \ldots \mu_i}_{\nu_1 \ldots \nu_i} \int d\omega\, d\varpi\, \mathcal{F}\{f_{\mu_1 \ldots \mu_i}\}(\omega) \mathcal{F}\{f^{\nu_1 \ldots \nu_i}\}(\varpi) \delta(\omega + \varpi) \\ &= \Delta^{\mu_1 \ldots \mu_i}_{\nu_1 \ldots \nu_i} \int d\omega\, \omega_{\mu_1} \cdot \ldots \cdot \omega_{\mu_i} \cdot \omega^{\nu_1} \cdot \ldots \cdot \omega^{\nu_i} \hat{f}(\omega) \hat{f}(-\omega) \end{aligned}$$

Because f is real, yielding a Fourier transform with even real part and odd imaginary part, we find

$$I = \Delta^{\mu_1 \ldots \mu_i}_{\nu_1 \ldots \nu_i} \int d\omega \omega_{\mu_1} \cdot \ldots \cdot \omega_{\mu_i} \cdot \omega^{\nu_1} \cdot \ldots \cdot \omega^{\nu_i} |\hat{f}(\omega)|^2$$

In this we can obviously interchange all indices under the integration without changing the result, which means that we can represent $\Delta^{\mu_1 \ldots \mu_i}_{\nu_1 \ldots \nu_i}$ as a symmetrized tensor. This symmetrized tensor can be represented on the form given in Lemma 2. Using Lemma 1 we can see that every term is on the same form, and we can add up $\sum_j a_j = \lambda$:

$$\sum_{j=0}^{[i/2]} a_j \int d\mathbf{x}\, \eta^{(\mu_1 \mu_2} \ldots \eta^{\mu_{2j-1} \mu_{2j}} \eta_{\nu_1 \nu_2} \ldots \eta_{\nu_{2j-1} \nu_{2j}} \delta^{\mu_{2j+1}}_{\nu_{2j+1}} \ldots \delta^{\mu_i)}_{\nu_i)} f_{\mu_1 \ldots \mu_i} f^{\nu_1 \ldots \nu_i} = \\ \lambda \int d\mathbf{x}\, \delta^{(\mu_1}_{(\nu_1} \ldots \delta^{\mu_i)}_{\nu_i)} f_{\mu_1 \ldots \mu_i} f^{\nu_1 \ldots \nu_i}$$

which is the same as (4).

□

This result might at first glanse seems a little surprising. The second order example is: the regularization using the squared trace of the Hessian as smoothness term yields the same result as if the trace of the squared Hessian was used. The reason that they yield the same result is that they are integrated over the spatial domain. We see as example:

$$\int_{x,y} dxdy\,(f_{xx}^2 + f_{yy}^2 + 2f_{xx}f_{yy}) = \int_{x,y} dxdy\,(f_{xx}^2 + f_{yy}^2 + 2f_{xy}^2)$$

which easily follows from partial integration or from the Fourier transform:

$$\int_{x,y} dxdy\,(f_{xx}^2 + f_{yy}^2 + 2f_{xx}f_{yy}) = \int_{\omega_1\omega_2} d\omega_1 d\omega_2\,(\omega_1^4 + \omega_2^4 + \omega_1^2\omega_2^2)|\hat{f}|^2 \quad .$$

This leads to the conclusion of this section:

Theorem 3 *The most general Cartesian invariant functional used for regularization in D dimensions is the functional*

$$E[f] = \int d\mathbf{x}\,(f-g)^2 + \sum_{i=1}^{N} \lambda_i f_{\mu_1\ldots\mu_i} f^{\mu_1\ldots\mu_i} \qquad (5)$$

having only one independent parameter λ_i per order of differentiation.

Proof We assume that the most general form of regularization functional is given in (3). In this case, the functional of N-th order regularization contains many degrees of freedom in the choice of the tensors $\Delta_{\nu_1\ldots\nu_i}^{\mu_1\ldots\mu_j}$. The choice is constrained be the requirements of invariance: $\Delta_{\nu_1\ldots\nu_i}^{\mu_1\ldots\mu_j}$ must be independent of f and x, i.e. it is a constant tensor. Furthermore, the functional should be an absolute Cartesian invariant, which means that $\Delta_{\nu_1\ldots\nu_i}^{\mu_1\ldots\mu_j}$ must be realizable as a polynomial in the metric tensor according to Theorem 1. Because we work in a Cartesian coordinate system, there is in principle no difference in upper and lower indices, and we will be able to bring the functional on a form, where $i = j$, by argumentation in the Fourier domain similar to that in Theorem 2. Then, according to Theorem 2, we can write the functional using only one parameter per order.

□

We have now proven, that even though the functional used for regularization contains all combinations of different derivatives of different orders, it only contains N independent constants, where N is the order of regularization, if the regularization is made in a rotationally (and translationally) invariant way. Using rotationally invariant regularization, we can easily generalize the results from the previous sections to higher dimensions. All that is needed is to substitute ω by $|\omega|$ in the filters. Also the semi-group property leads trivially to the Gaussian filter using the 1-norm addition. We can then state:

Result 3 *Rotationally invariant regularization having an equal weight of all orders of derivatives in a Taylor series sense leads to Gaussian scale space.*

4 Truncated smoothness operators

To be consistent with scale space Tikhonov regularization has to be of infinite order. Nevertheless, low-order regularization is often performed [10]. We might perceive low-order regularization as a regularization using a truncated Taylor-series of the smoothness functional. Here we list the linear filters h_n found by Fourier inversion. The notation h_n indicates the filter where all λ_i are zero except those mentioned by their index in the set n.

$$h_1(x) = \frac{1}{\sqrt{2\lambda}} e^{\frac{-|x|}{\sqrt{\lambda}}} \tag{6}$$

$$h_2(x) = \frac{\pi}{\lambda^{1/4}} \cos\left(\frac{\sqrt{2}|x|}{\lambda^{1/4}}\right) e^{-\frac{|x|}{\sqrt{2}\lambda^{1/4}}} \tag{7}$$

In the case of mixed first and second order regularization with $\lambda = \lambda_1$ and $\lambda_2 = \lambda_1^2/2$ corresponding to the truncation to second order of the energy functional which implies the Gaussian filtering, we find

$$h_{12}(x) = \frac{\pi(|x| + \sqrt{\lambda})}{2\lambda} e^{-\frac{|x|}{\sqrt{\lambda}}}$$

We notice, that the first order regularization filter is always positive, while the second order filter is similar apart from a multiplication with an oscillating term. The latter explains why second order regularization might give inexpedient and oscillating results (so-called overshooting or ringing effects). The second order truncated Gaussian filter h_{12} is always positive, and will in general give more sensible results than the one without the first order term. We will now look into how in general to construct regularization filters, which are not oscillating in the spatial domain.

Proposition 4 *When the regularization constants λ_i have values so that the polynomial*

$$p(x) = 1 + \lambda_1 x + \lambda_2 x^2 + \lambda_3 x^3 \ldots \lambda_N x^N$$

have roots all real and negative, then the regularization filter

$$\hat{h}(\omega) = \frac{1}{1 + \sum_{i=1}^{\infty} \lambda_i \omega^{2i}}$$

is positive for all $x \in \mathbb{R}$ in the spatial domain.

Proof When writing $p(x)$ on factorial form

$$p(x) = \Pi_i (1 + a_i x)$$

the a_is are minus the roots and thereby all positive and real. The regularization using λ_i results in the filter in the Fourier domain

$$\hat{h}(\omega) = \frac{1}{p(\omega^2)} = \frac{1}{\Pi_i (1 + a_i \omega^2)}$$

with all a_i positive and real. In the spatial domain this is the convolution product of $h_1(x)$, which is a positive filter, with itself (in re-scaled versions). This yields a positive filter.

$\square$

As an example we can construct a nth order regularization filter which is always positive in the spatial domain. Let all $a_i = s/n$. We find the regularization parameters

$$\lambda_i = K_i^n \frac{s}{n}$$

where K_i^n is the number of possible i-sets taken out of n. We notice when n is increased towards infinity, that we find the Gaussian:

$$\lim_{n\to\infty} - \log \hat{h} = \lim_{n\to\infty} \sum_{i=1}^{n} \log(1 + \frac{s}{n}\omega^2) = s\omega^2$$

Also if the a_i had taken different values, the resulting filter would after renormalization be a Gaussian. We conclude:

Result 4 *The only infinite order regularization filter which can be decomposed into first order regularization filters with positive and real λs is the Gaussian.*

We find the other semi-group filters, by substitution of the first order regularization by n order regularization, where $\lambda_i = 0$ for all $i < n$ in all the regularization filters.

The truncated or low order regularization filters can be efficiently implemented by recursive filtering [12]. This is due to the structure of being a reciprocal polynomial with only even orders of the frequency. In section 6, we show, that any regularization can be implemented as a recursive filtering using no more than $2n$ multiplications and additions per output element.

So far we have introdused two different ways of approximating a Gaussian filter: using the truncated Taylor series expansion or using the positive filter of finite order with equal a_is. Non of these approximations are claimed to be optimal. The first will have ringing effects due to the oscillating filters. The second is not necessarily, the closest one can get to a Gaussian given the order of regularization. Deriche [14] has proposed several methods to optimally approximate the Gaussian by low order recursively filtering. These filters are oscillating, but in a least squares sense closest to the Gaussian.

The second order truncated Gaussian h_{12} is the filter that was proposed by Deriche [8] as an optimal smoothing operator. The filter simultaneously maximizes the criteria proposed by Canny [7]: Signal-to-noise-ratio, localization and uniqueness of the zero-crossings of the second derivative of the filter answer of a noise-corrupted step edge. The filter have been efficiently implemented by recursive filtering [12] using only 8 multiplications and 7 additions for each output element. In the following we will look into the Canny criteria of optimality and propose generalizations.

5 Canny optimality

Canny proposed criteria of optimality of a feature detection filter. The feature is detected where the maxima of the linear filtering occur. In the case of a general feature $e(x)$ with uncorrelated Gaussian noise the measures of signal-to-noise ratio Σ, localization Λ, and uniqueness of zero-crossing Υ are:

$$\begin{aligned}\Sigma &= \frac{\left|\int e(x)f(x)dx\right|}{\left(\int f^2(x)dx\right)^{1/2}}\\ \Lambda &= \frac{\left|\int e'(x)f'(x)dx\right|}{\left(\int f'^2(x)dx\right)^{1/2}}\\ \Upsilon &= \frac{\left(\int f'^2(x)dx\right)^{1/2}}{\left(\int f''^2(x)dx\right)^{1/2}}\end{aligned}$$

where all integrals are taken over the real axis. We now define the feature to be a symmetric step edge:

$$e(x) = \int_{-\infty}^{x} \delta_0(t)dt$$

The symbol δ_0 denotes the Dirac delta function. Canny tries first simultaneously to maximize Σ and Λ and finds the box filter as the optimal solution on a finite domain. To avoid the box-filter he then introduces the uniqueness measure. After this, a simultaneously optimization of all three measures using Lagrange multipliers is performed. Deriche [8] finds the optimal solution on a infinite domain. In the following we will show, that the uniqueness criteria can be omitted on the infinite domain leading to a conceptually simpler result, but with the lack of simplicity of the edges.

In order to find the optimal smoothing filter h which can be differentiated so as to give the optimal step edge detector, we substitute $h'(x) = f(x)$ in the measures and try to find h. We use one of the factors as optimality criterion and the others as constraints to find a composite functional Ψ[3] (using Lagrange multipliers). To obtain symmetry in this formulation we multiply all factors

[3]At this point, we might notice the resemblance of the composite functional and the energy functional in regularization.

by an arbitrary Lagrange multiplier λ_i.

$$\Psi[h] = \lambda_1 e h' + \lambda_2 h'^2 + \lambda_3 e' h'' + \lambda_4 h''^2 + \lambda_5 h'''^2$$

Omitting the uniqueness constraint corresponds to $\lambda_5 = 0$. A necessary condition to have an optimal filter is that the first variation of this is zero

$$\frac{\delta\Psi}{\delta\varepsilon} = -\lambda_1\delta_0 - 2\lambda_2 h^{(2)} + \lambda_3\delta_0^{(2)} + 2\lambda_4 h^{(4)} - 2\lambda_5 h^{(6)} = 0$$

Here, a parenthesized superscript denotes order of differentiation. By Fourier transform of this we find

$$\hat{h}(\omega) = \frac{\lambda_1 + \lambda_3\omega^2}{2(\lambda_2\omega^2 + \lambda_4\omega^4 + \lambda_5\omega^6)}$$

The integral over the filter in the spatial domain must be 1. This means that $\lambda_1 = 0$ and $\lambda_3 = 2\lambda_2$ and we find

$$\hat{h}(\omega) = \frac{1}{1 + \alpha\omega^2 + \beta\omega^4}$$

where $\alpha = 2\lambda_4/\lambda_3$ and $\beta = 2\lambda_5/\lambda_3$. Deriche chooses $\alpha = 2\sqrt{\beta}$, and fixes in this way a one parameter family of optimal filters. The choice corresponds well with a dimensional analysis. Another way to select a one parameter family is to set $\beta = 0$. This corresponds to omitting the uniqueness criteria. Setting $\lambda_1 = 0$ means in principle, that we do not take signal value into account in the signal-to-noise ratio. This is not a conceptual problem, because our demand of having a normalized and symmetric filter automatically fixes the signal value in the case of a step edge.

By omitting the uniqueness criteria we find the optimal step edge detection filter to be the derivative of

$$\hat{h}(\omega) = \frac{1}{1 + \alpha\omega^2}$$

This is exactly the same as in the first order regularization. It seems intuitively correct that the optimal first order filter should project into a Sobolev space of first order. We do not in general want the second derivative of the signal to exist, when we only look at first order characteristics. In spatial domain, we find the filter

$$h'(x) = \frac{\pi}{\sqrt{\alpha}} \frac{x}{|x|} e^{-|x|/\sqrt{\alpha}}$$

The derivatives of this filter are not well-defined, and we see, that we indeed have projected into a Sobolev space of first order. When the derivatives of the filters are not well-defined, the filtered image, will not fullfil regularity criteria: It is not differentiable, etc. This means that the structure of the edge image is not necessarily simple. The uniqueness criteria (or the criteria of the edge image to be differentiable) has to be added to have a simpler structure.

5.1 A higher order optimal filter

We now define a nth order step edge e_n to be the function where all derivatives of order exceeding n are zero, the derivative of order n is a step edge function, and all lower order derivatives are continuous.

Definition 3 *A nth order step edge is*

$$e_n = \begin{cases} -k_n x^n & \text{for } x < 0 \\ k_n x^n & \text{for } x \geq 0 \end{cases}$$

where k_n is the strength of the edge.

We now want to find the smoothness filter which is simultaneously optimal for all orders up to n, in the sense that the first derivative of the filter is optimal to detect zero order step edges, the second derivative of the filter is optimal to detect first order step edges and so forth.

Proposition 5 *A nth order optimal step edge detector $h_n(x)$ can be written on the form*

$$\hat{h}_n = \frac{1}{\sum_{i=0}^{n}(\gamma'_i + \alpha'\gamma'_{i-1} + \beta'\gamma'_{i-2})(\tau\omega^2)^i}$$

where γ'_i is the dimensionless part of the weight of the ith order optimality, α' and β' are the dimensionless part of the weighing of the different parts of the optimality criteria. $\gamma'_i \equiv 0$ for all $i < 0$.

Proof We notice that we can write the nth order step edge on the form

$$e_n = \frac{k_n \delta_0^{-(n+1)}}{n!}$$

where $\delta_0^{-(n+1)}$ denotes the Dirac delta function integrated $n+1$ times.

We now find the composite functional of order n Ψ^n by adding the factors arising from h' being an optimal zero order step edge detector, h'' being an optimal first order detector and so forth, up to an nth order edge. In this way a smoothing function, which is simultaneously optimal in several orders is constructed. We weigh the ith order optimality by γ_i and find

$$\begin{aligned}\Psi^n &= \sum_{i=0}^{n} \lambda_1 \gamma_i \frac{k_i}{i!} \delta_0^{(-i-1)} h^{(i+1)} + \lambda_2 \gamma_i (h^{(i+1)})^2 \\ &\quad + \lambda_3 \frac{k_i}{i!} \delta_0^{(-i)} h^(i+2) \gamma_i + \lambda_4 \gamma_i (h^{(i+2)})^2 + \lambda_5 \gamma_i (h^{(i+3)})^2\end{aligned}$$

The first variation of this yields

$$\frac{\delta \Psi^n}{\delta \varepsilon} = \sum_{i=0}^{n} \lambda_1 \frac{k_i}{i!} \delta_0 \gamma_i + \lambda_2 \gamma_i h^{(2i+2)} + \lambda_3 \frac{k_i}{i!} \delta'' \gamma_i + \lambda_4 \gamma_i h^{(2i+4)} + \lambda_5 \gamma_i h^{(2i+6)}$$

This must be zero and we find by Fourier transformation (after noticing that $\lambda_1 = 0$ in order to make the filter normalizable in the spatial domain) the following optimal filter:

$$\begin{aligned}\hat{h} &= \frac{\sum_{i=0}^{n} \frac{k_i}{i!} \gamma_i \lambda_3}{\sum_{i=0}^{n} \gamma_i \lambda_2 \omega^{2i} + \gamma_i \lambda_4 \omega^{2i+2} + \gamma_i \lambda_5 \omega^{2i+4}} \\ &= \frac{\sum_{i=0}^{n} \frac{k_i}{i!} \gamma_i \lambda_3}{(\lambda_2 + \lambda_4 \omega^2 + \lambda_5 \omega^4) \sum_{i=0}^{n} \gamma_i \omega^{2i}}\end{aligned}$$

We assume that zero order edges have non-zero weight and write without loss of further generality

$$\hat{h} = \frac{1}{(1 + \alpha \omega^2 + \beta \omega^4) \sum_{i=0}^{n} \gamma_i \omega^{2i}}$$

where we have defined $\gamma_0 \equiv 1$ and used the normalization in the spatial domain. This is the general form of a up-to-nth-order-Canny-optimal filter. If we gather terms of equal power of ω, we find

$$\hat{h} = \frac{1}{\sum_{i=0}^{n}(\gamma_i + \alpha\gamma_{i-1} + \beta\gamma_{i-2})\omega^{2i}}$$

where γ with negative index is defined to be 0. A dimensional analysis shows

$$[\gamma_i] = [\omega^{-1}]^{2i} \quad [\alpha] = [\omega^{-1}]^2 \quad [\beta] = [\omega^{-1}]^4$$

We write the constants as products of a dimensionless constant and a dimensionful term expressed as powers of τ, where $[\tau] = [\omega^{-1}]^2$:

$$\hat{h} = \frac{1}{\sum_{i=0}^{n}(\gamma_i' + \alpha'\gamma_{i-1}' + \beta'\gamma_{i-2}')(\tau\omega^2)^i}$$

□

This filter *can* be a Taylor series expansion of the Gaussian if the γs are chosen carefully. The choice of γs depends in this case on α' and β'.

If we choose all orders (up to infinity) of step edges to be equally important in a Taylor series sense, we find the smoothing filter

$$\hat{h} = \frac{e^{-\tau\omega^2}}{1 + \alpha'\tau\omega^2 + \beta'\tau^2\omega^4}$$

which we might perceive in two different ways: it is the Deriche smoothing filter, which has been convolved with a Gaussian in order to expand it to infinite order (make it project into a Sobolev space of infinite order), or it is a Gaussian which has been low pass filtered in order to make it Canny-optimal. When we make the choice $\alpha'^2 = 2\sqrt{\beta'} = 1$, we find in the spatial domain the filter:

$$h(x,\tau) = \frac{\pi}{8\tau^2}(p(x,\tau) + p(-x,\tau))$$

where

$$p(x,\tau) = \sqrt{2}e^{\sqrt{x}\tau + \frac{1}{2}} \cdot x \cdot (1 + \mathrm{erf}(\frac{1}{\sqrt{2}}(1 + \frac{x}{\tau}))) + \frac{2\tau}{\sqrt{\pi}}e^{-\frac{x^2}{2\tau^2}}$$

and erf() is the error function.

By omitting the uniqueness term, we find the up to nth order optimal filter

$$\hat{h} = \frac{1}{\sum_{i=0}^{n}(\gamma_i' + \alpha'\gamma_{i-1}')(\tau\omega^2)^i}$$

This filter projects into a Sobolev space of order n, and the solution corresponds to nth order regularization.

6 Implementation issues

The regularization using quadratic stabilizers can be implemented in various ways. Firstly the energy minimization can be implemented by a gradient descend algorithm (or some orther minimization algorithm). This has the advantagethat the natural boundary conditions can be implemented directly, leading to sensible results near the boudaries. The disadvantage is slow convergence and thereby long computational time. Especially, for large λs this will be time consuming.

The regularization can be implemented as a convolution in the spatial domain (in the cases, where the analytic expression of the filter is known in the spatial domain). In this case the boundaries can be handled by cutting off the filters and renormalization. The computational complexity will be $O(MN\lambda^2)$, where the size of the image is $M \times N$ pixels, for a first order regularization.

The filtering can also be implented in the Fourier domain as a multiplication, using the Fast Fourier Transform (FFT). In this case the image is assumed to be cyclic, which might imply strange phenomena near the boundaries. The computational complexity is $O(MN \log M \log N)$ indepedently of the order of regularization and the λs. In order to make the cyclic boundary effects smaller, one can embed the image in a $K \times L$ image, but this will enlarge the constant in the complexity measure.

Finally, the regularization can be implemented as a recursive filtering. In this case, the boundaries can be handled by cutting off like in the case of convolution in the spatial domain. The computational complexity is $O(MNn)$, where n is the order of regularization. In most practical cases, this will be the fastest implementation. This is why we show how to do in practice:

6.1 Recursive implementation

In order to deal with the recursive system, we need to reformulate the regularization on a discrete grid. We define the energy

$$E(f) \equiv \sum_x \left((f-g)^2 + \sum_{i=1}^{N} \lambda_i (d_i * g)^2 \right)$$

where d_i is the ith order difference filter. We recall the definition of the z-transform

$$\hat{h} = \sum_{x=-\infty}^{\infty} h(x) z^{-x}$$

and make the following proposition:

Proposition 6 *The discrete regularization can be implemented by recursive filtering using no more than* $2N$ *multiplications and additions per output element, where* N *is the order of regularization.*

Proof By following the argumentation in the continuous case (and substituting the Fourier transform by the z-transform) or the discrete formulation given by Unser et. al. [13], we find that the minimization is implemented by convolution with the filter

$$\hat{h}(z) = \frac{1}{1 + \sum_{i=1}^{N} \lambda_i \hat{d}_i(z) \hat{d}_i(z^{-1})}$$

where the hat indicates the z-transform. The transform of the difference operator is given by

$$\hat{d}_i(z) = z^{-i/2}(1-z)^i$$

which implies that $\hat{d}_i(z)\hat{d}_i(z^{-1})$ is a N order polynomion in z multiplied by a N order polynomion in z^{-1}. Because the transfer function is symmetric in z and z^{-1} the $2N$ roots in the denominator will appear in pairs of z and z^{-1}. From this follows, that the transfer function can be decomposed as

$$\hat{h}(z) = \hat{h}^{+}(z)\hat{h}^{+}(z^{-1})$$

where
$$\hat{h}^+(z) = \frac{c}{\prod_{i=1}^{N}(1 - z_i z^{-1})} \equiv \frac{k}{1 - a_i z^{-1} - \ldots - a_N z^{-N}}$$
where z_i are the roots in the denominator ordered by length. This means that the regularization can be implemented by the forward and backwards recursive filters with identical coefficients:

$$\begin{aligned} f^+(x) &= g(x) + a_1 f^+(x-1) + \ldots + a_N f^+(x-N) \\ f^-(x) &= f^+(x) + a_1 f^-(x+1) + \ldots + a_N f^-(x+N) \\ f(x) &= k^2 f^-(x) \end{aligned}$$

□

We have here given the proof, that regularization in the discrete case can be implemented (without any other approximation than the discrete implementation) as recursive filtering. The proof follows the lines of the proof given by Unser et. al. [13] for the case, where only a single stabilizer is used.

7 Conclusion

We have presented a formulation of the Heat Equation which corresponds to a Thikonov regularization and is formulated as a minimization of a functional. Furthermore, we have developed an infinite series of regularization filters, which are all having the semi group property. The simplest of these is the well-known Gaussian. The series of semi-group filters converges towards an ideal low-pass filtering, when the scale addition convention converges towards the infinity norm.

The minimization formulation of Gaussian scale space leads to a set of boundary conditions, which is consistent, and does not imply any assumption of periodicity or truncation of filters. Furthermore, as the boundary conditions are consistent with the Heat Equation defined on a finite domain, the causality criterion is fulfilled. This is not necessarily the case for truncated filters.

We have shown, that the Canny-Deriche filter is a second order regularization filter, when we choose coefficients corresponding to the truncation of the functional leading to Gaussian scale space. The Canny-Deriche filter is optimized for detection of zero order step edges. We have shown that the criterion of uniqueness is not necessary, to well-define the filter. We have generalized the Canny-Deriche filter to any order and have shown, that the filter which is simultaneously optimized for all orders up to n corresponds to an $(n+1)$th order regularization. If the uniqueness is omitted it corresponds to a nth order regularization.

Furthermore, we have shown that regularization in a higher dimensional space in its most general form is a rotation of the 1D regularization, when Cartesian invariance is imposed. This means that all of the above result, can be straight forward generalized to any dimension, under the condition, that the functional is a Cartesian invariant.

In the following, we show some examples of filterings using the different regularization filters. In all examples, the filtered image, level sets of the filtered image, the horizontal derivative of the filtered image, and level sets of the derivative are shown. The effects of non-regularity of the first derivative of a first order regularized image, the ringing of a second order regularizing image etc. are visible. In all cases, we have used the Cartesian invariant smoothing kernel, and its derivatives. The same thresholds have been used to create the level sets in all images, which is why they are immediatly comparable.

The above work only dealt with linear problems: Linear convolution, linear scale space, linear biased diffusion. In future a major subject for studies will be to find the resemblance between non-linear biased and non-biased diffusion. Can we transform a non-linear diffusion equation [15] into a non-linear biased diffusion equation [16] (and thereby capture minimization of non-quadratic functionals) and visa versa?

8 Acknowledgements

We would like to thank Tony Lindeberg for valueable discussions, and Robert Stahr for his carefully and patiently adaption of his programs to our purpose.

References

[1] L. Schwartz, "*Theorie des Distributions*", vol. I, II of Actualités scientifigues et industrielles, Publications de l'Institut de Mathématique de l'Université de Strasbourg, 1950-51.

[2] A. P. Witkin, "*Scale Space Filtering*", Proc. of IJCAI, Karlsruhe, Germany, 1983.

[3] J. J. Koenderink, "*The Structure of Images*", Biol. Cybern., vol. 50, pp. 363-370, 1984.

[4] L. Florack, "*The Syntactical Structure of Scalar Images*", PhD-thesis, University of Utrecht, 1993.

[5] A. N. Tikhonov and V. Y. Arseninn, "*Solution of ill-posed problems*" V.H Winston & Sons, John Wiley, 1977.

[6] T. Poggio, H. Voorhees and A. Yuille, "*A Regularized Solution to Edge Detection*", A. I. Memo 883, M.I.T, May 1985.

[7] J. F. Canny, "*A Computational Approach to Edge Detection*", IEEE PAMI, Vol. 8, No. 6, Nov 1986.

[8] R. Deriche, "*Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector*", Int. Jour. of Computer Vision, pp. 167-187, 1987.

[9] T. Lindeberg, "*Scale-Space Theory in Computer Vision*", Kluwer Academic Publishers, Netherlands, 1994.

[10] W. E. L. Grimson, "*From images to surfaces*", MIT Press, 1981.

[11] E. J. Pauwels, P. Fiddelaers, T. Moons, and L. J. V. Gool, "*An extended Class os Scale-Invariant and Recursive Scale Space Filters*", Katholieke Universiteit Leuven, Tech. Rep. KUL/ESAT/MI2/9316, 1994.

[12] R. Deriche, "*Fast Algorithms for Low-level Vision*", IEEE PAMI, Vol. 12, No. 1, Jan 1990.

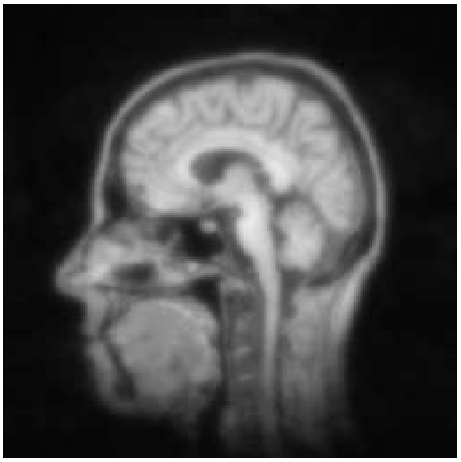
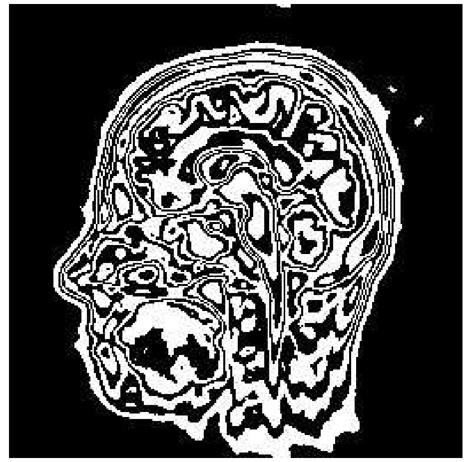
[13] M. Unser, A. Aldroubi, and M. Eden, "*Recursive Regularization Filters: Design, Properties and Applications*", IEEE PAMI, vol. 13, no. 3, March 1991.

[14] R. Deriche "*Recursively implementing the Gaussian and its derivatives*", INRIA, Tech. Rep. 1893, May 1993.

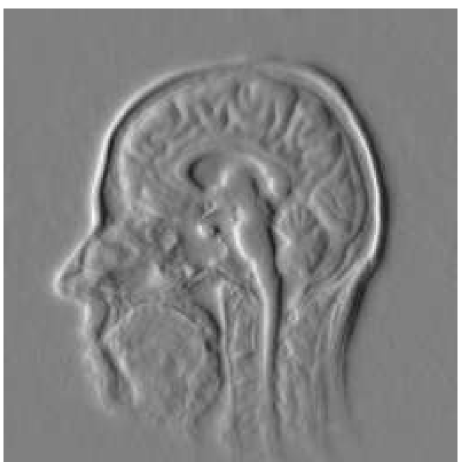
[15] P. Perona and J. Malik, "*Scale Space and Edge Detection using Anisotropic Diffusion*" IEEE PAMI, vol. 12, pp. 429-439, 1990.

[16] N. Nordström, "*Biased Anisotropic Diffusion - a Unified Regularization and Diffusion Approach to Edge Detection*", Image and Visual Computing, no. 8, pp. 318-327, 1990.

Figure 1: Brain image which is used in some examples.

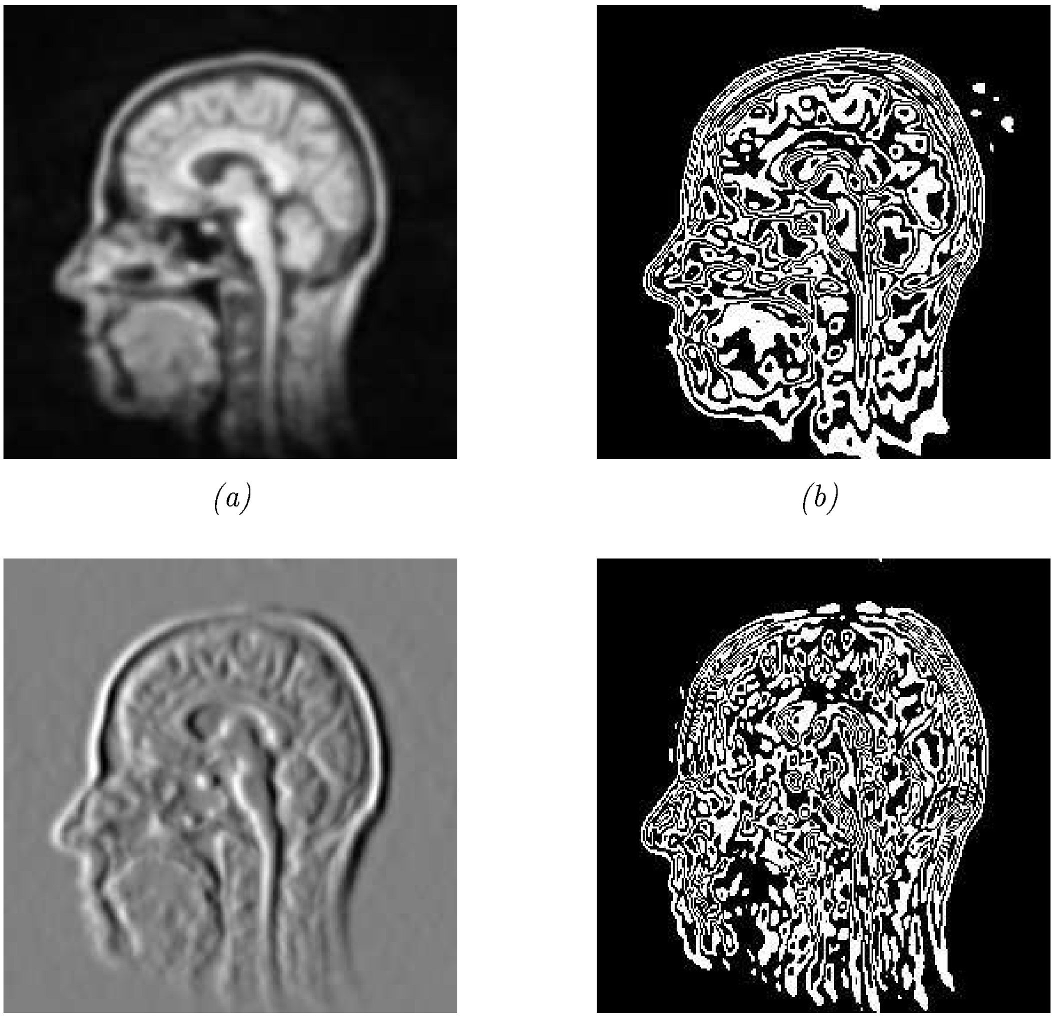
 

(a) *(b)*

(c) *(d)*

Figure 2: First order regularized brain ($h(\omega) = \frac{1}{1+\lambda_1\omega^2}$). (a) is the filtered image ($\lambda_1 = 6$ pixels), (b) is level sets of (a), generated by dividing the itensities into intervals and letting every second interval become black and every second white, (c) is the horizontal first derivative of (a), and (d) is level sets of (c). Notice that the first derivative is not regular in theory. It is contained in a Sobolev space of order 0, but not necessarily in one of order 1.



(a) (b) (c) (d)

Figure 3: Second order regularized brain ($h(\omega) = \frac{1}{1+\lambda_2\omega^4}$). ($a$) is the filtered image ($2\sqrt{\lambda_2} = 6$ pixels), (b) is level sets of (a), (c) is the horizontal first derivative of (a), and (d) is level sets of (c). Notice the ringing effects.

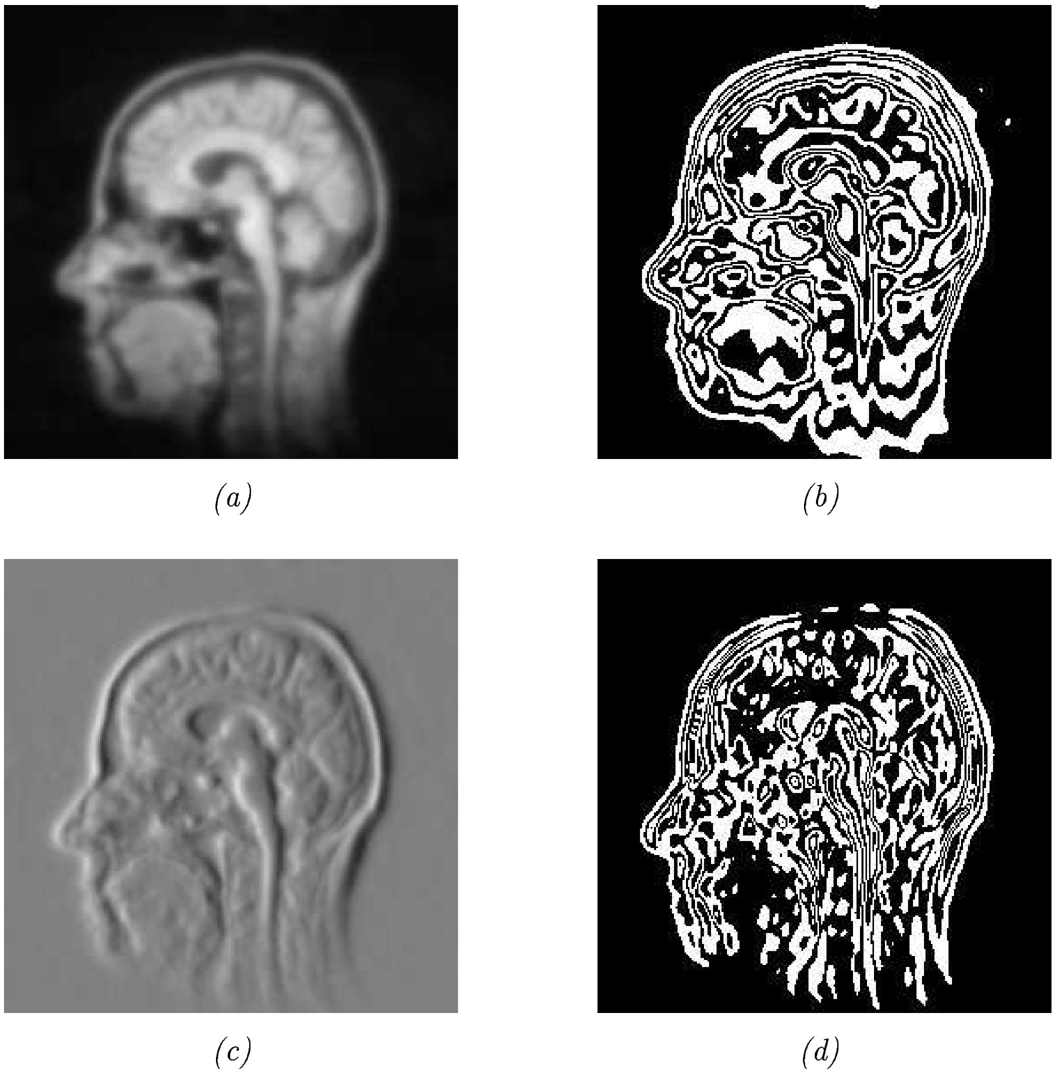
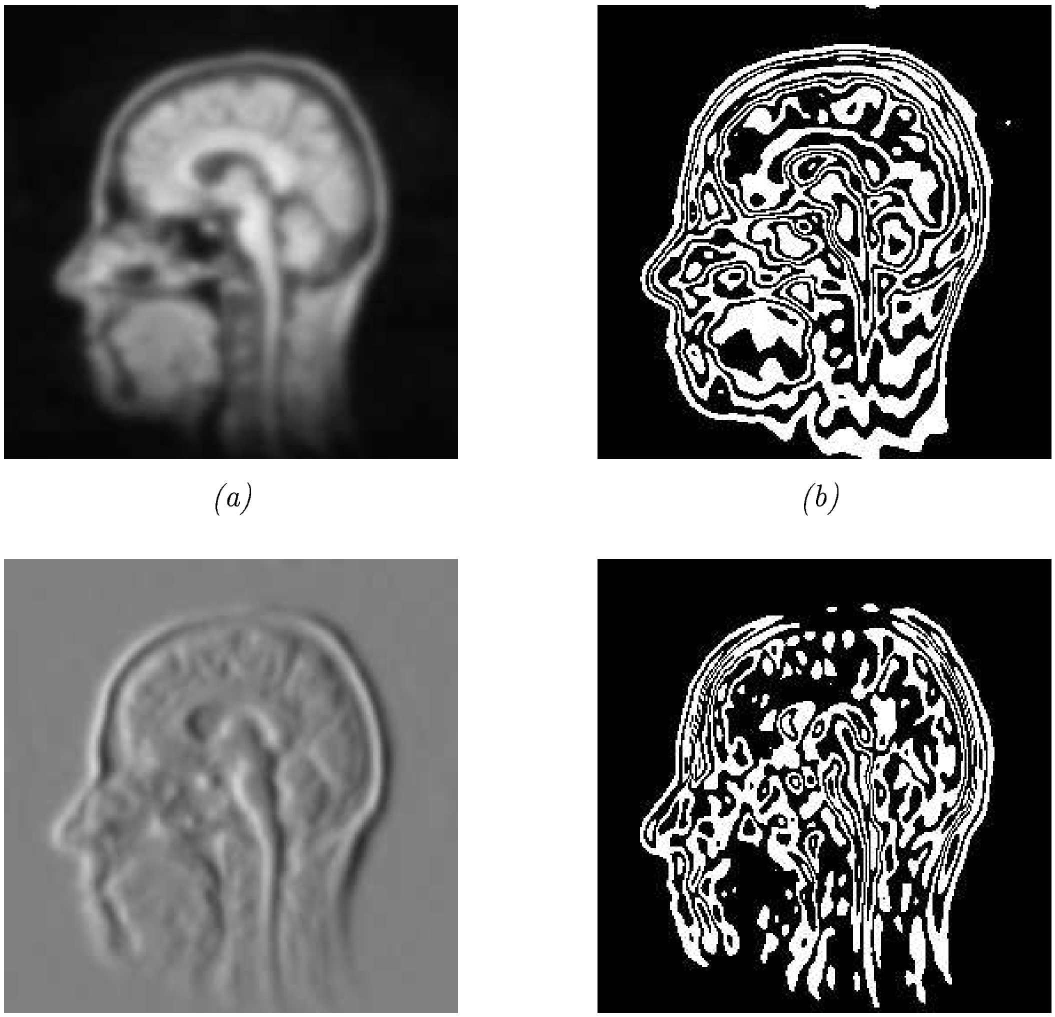


Figure 4: First and second order (truncated Gaussian) regularized brain $(h(\omega) = \frac{1}{1+\lambda_1\omega^2+\lambda_2\omega^4})$. (a) is the filtered image ($\lambda_1 = 2\sqrt{\lambda_2} = 6$ pixels), (b) is level sets of (a), (c) is the horizontal first derivative of (a), and (d) is level sets of (c). Notice the regularity of the derivative compared to the derivative of the first order regularized brain in Figure 2. It is contained in a Sobolev space of order1.



(a) (b)

(c) (d)

Figure 5: Fourth order truncated Gaussian regularized brain ($h(\omega) = \frac{1}{(1+\lambda_1\omega^2)^4}$). The filter is always positive in the spatial domain, but does not correspond to the truncation of the Taylor series of the Gaussian. (a) is the filtered image ($\lambda_1 = 6$ pixels), (b) is level sets of (a), (c) is the horizontal first derivative of (a), and (d) is level sets of (c).

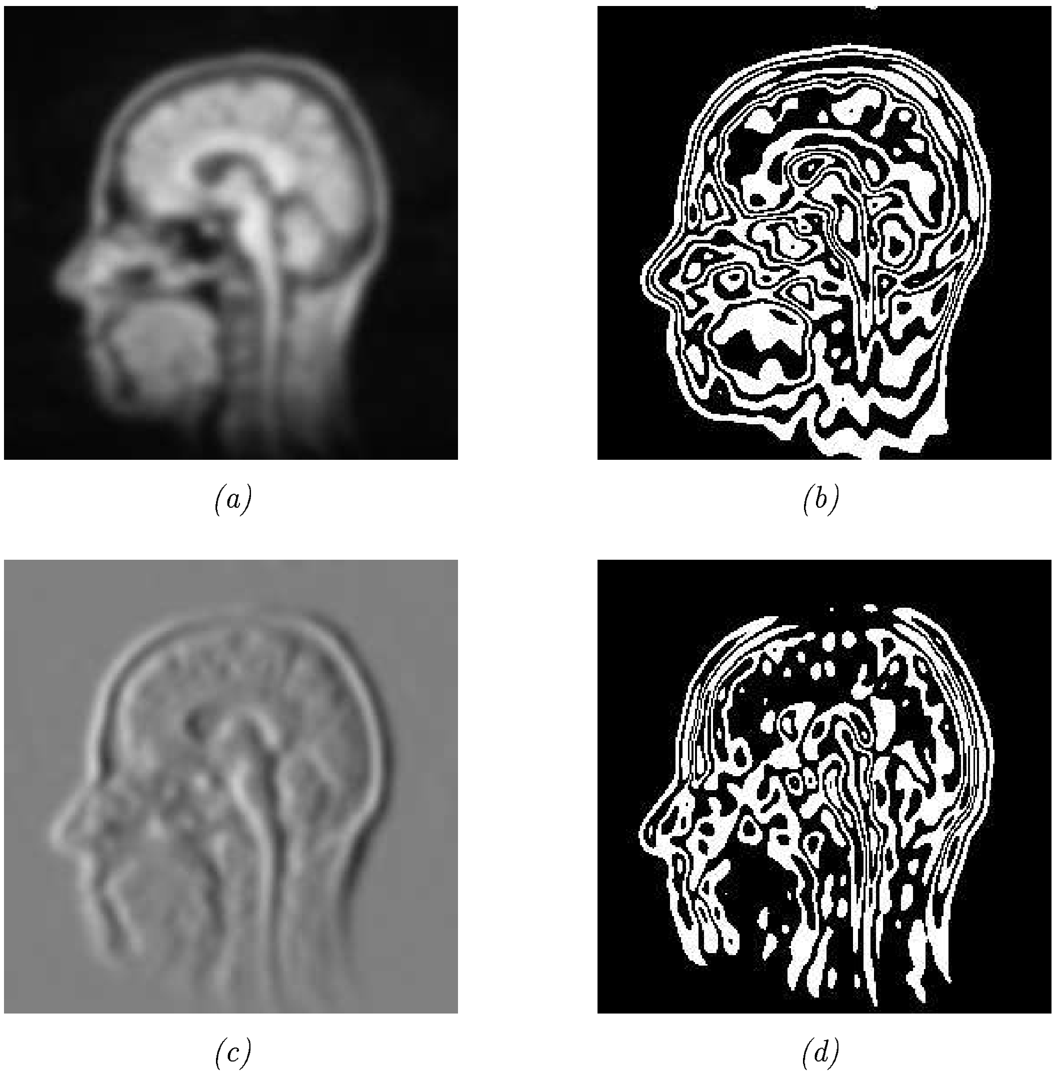
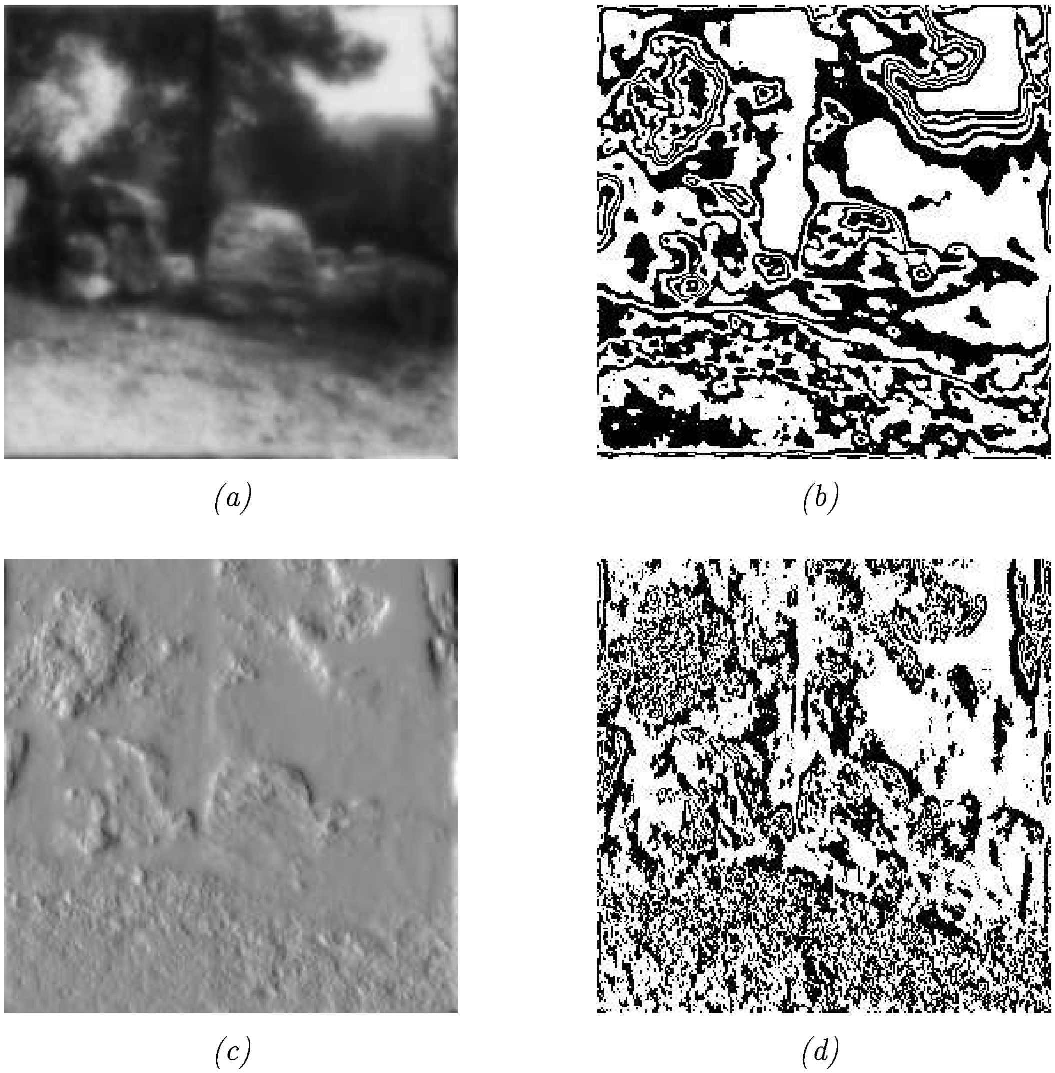


Figure 6: Gaussian filtered brain ($h(\omega) = exp{-}\omega^2\lambda_1$) . ($a$) is the filtered image (σ corresponding to $\lambda_1 = 6$ pixels), (b) is level sets of (a), (c) is the horizontal first derivative of (a), and (d) is level sets of (c).

Figure 7: Image of outdoor scene which is used in the following examples.



(a) (b) (c) (d)

Figure 8: First order regularizedoutdoor scene ($h(\omega) = \frac{1}{1+\lambda_1\omega^2}$). (a) is the filtered image ($\lambda_1 = 6$ pixels), (b) is level sets of (a), generated by dividing the itensities into intervals and letting every second interval become black and every second white, (c) is the horizontal first derivative of (a), and (d) is level sets of (c). Notice that the first derivative is not regular in theory. It is contained in a Sobolev space of order 0, but not necessarily in one of order 1.

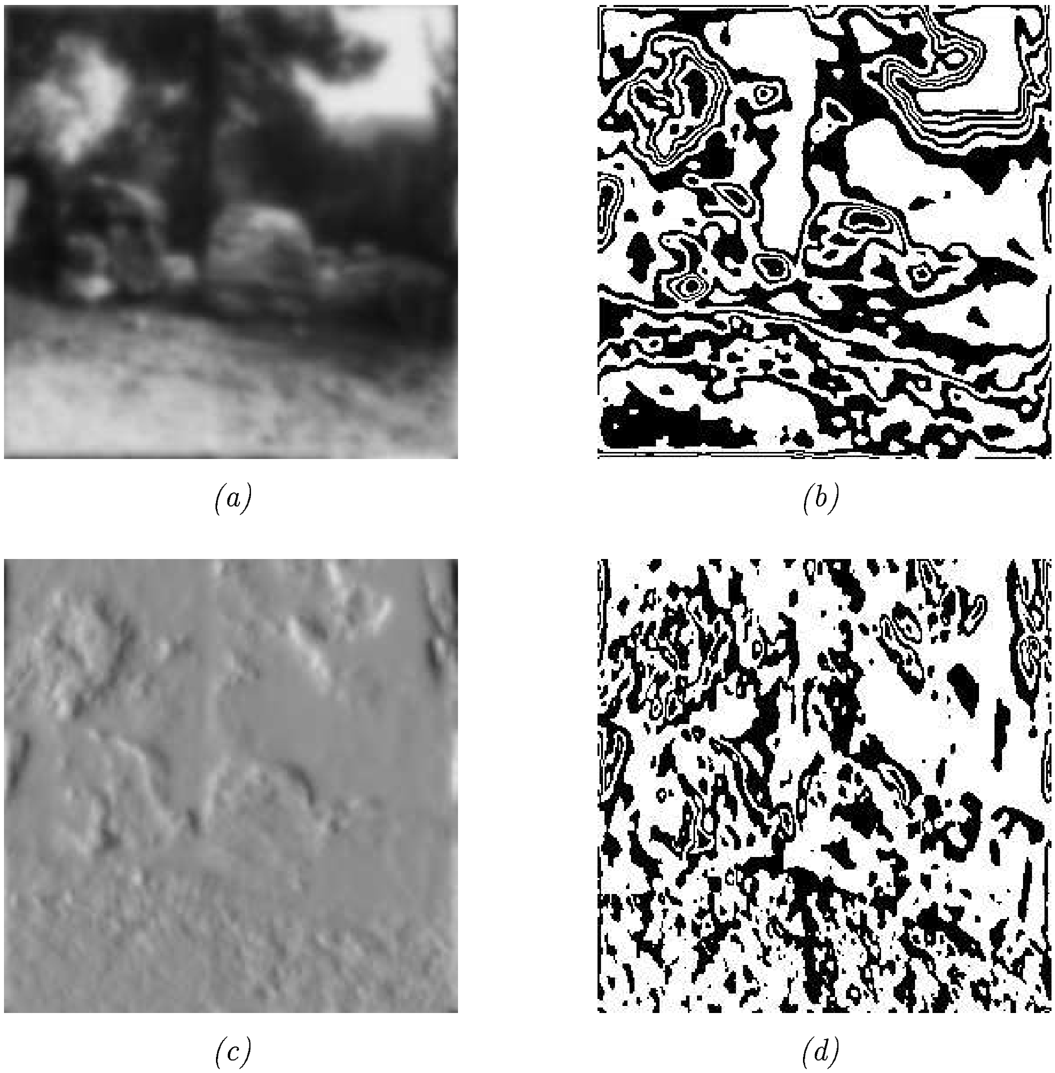


Figure 9: First and second order (truncated Gaussian) regularized outdoor scene ($h(\omega) = \frac{1}{1+\lambda_1\omega^2+\lambda_2\omega^4}$) . (a) is the filtered image ($\lambda_1 = 2\sqrt{\lambda_2} = 6$ pixels), (b) is level sets of (a), (c) is the horizontal first derivative of (a), and (d) is level sets of (c). Notice the regularity of the derivative compared to the derivative of the first order regularized brain in Figure 8. It is contained in a Sobolev space of order 1.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399